
Texas Open Data Portal
Resource Guide | May 2020

Texas Open Data Portal Resource Guide

Office of the Chief Data Officer for the State of Texas

Texas Department of Information Resources

First published as the *Open Data Playbook* in June 2017 by the
Texas Department of Information Resources
300 W. 15th Street, Suite 1300, Austin, Texas 78701

Revised in May 2020

Contents

Welcome from the Texas Data Management Program 1

Acknowledgements 2

1. Open Data Overview 3

1.1 What is Open Data? 3

Excluded Data 3

1.2 The Value of Open Data 4

Reduction of Public Information Requests 5

Self-Service Model 5

1.3 The Open Data Lifecycle 6

2. Open Data Portal 7

2.1 Open Data Portal 7

Background 7

Differences with the Closed Data Portal 7

2.2 Roles and Responsibilities 8

DIR Open Data Portal Team 8

Open Data Portal Users Group (ODPUG) 8

Socrata 8

Agencies 8

3. Data Governance 10

3.1 Establishing an Open Data Governance Framework 10

Open Data Stewards 10

3.2 Creating an Open Data Inventory 11

What is an Open Data Inventory? 11

Open Data Inventory and Metadata 12

4. Data Preparation 13

4.1 Data Assessment and Selection 13

Guiding Questions for Data Assessment 13

4.2 Data Quality 14

Data Quality Checkpoints 14

4.3 Data Privacy 15

Guiding Questions for Data Privacy 15
5. Publishing Open Data 16
5.1 Creating Metadata for a Dataset 16
Dataset Level Metadata 16
Column Level Metadata 17
5.2 Manual Publishing 18
5.3 Automated Publishing 18
DataSync 18
Socrata Open Data API 19
5.4 Publishing Process 20
Private vs. Public Data 20
6. Data Consumption 22
6.1 Filtering and Sorting 22
6.2 Charts, Graphs, and Maps 22
Data Lens 23
6.3 Data Stories 24
7. Checklist for Getting Started on the Open Data Portal 26
8. Socrata Open Data Portal Education Plan 27
Recommended Course Order for New Users 27
Foundational Courses 27
9. Additional Resources and Links 28
Socrata Support 28
External Open Data Portals 28
10. Glossary of Terms 29
11. Appendix: Open Data Legislation in Texas 30

Welcome from the Texas Data Management Program

Open data is a critical component of a trusted and reliable government. Constituents look to their government to provide transparency of business and operational practices. This transparency allows constituents to hold their public officials accountable to serve in the best interests of the public good.

In 2018, the 86(R) Texas Legislature introduced SB 819 which focused language on the continued importance of open data and the Texas Open Data Portal and on September 1st, 2019, the bill became law. SB 819 states that the Department of Information Resources (DIR) "shall establish a central repository of publicly accessible electronic data as the official open data Internet website for this state" and that the repository shall be designated as the Texas Open Data Portal (ODP).

The Office of the Chief Data Officer recognizes the importance and value of the ODP and provides state agencies and political subdivisions shared access to the repository to post publicly accessible open data. In leveraging the ODP, and working together, we can ensure that the business of our government is transparent and performing up to the expectations of those whom we serve.

Sincerely,

Ed Kelly

Chief Data Officer
Texas Department of Information Resources

One of the five key service areas of the Texas Data Management Program (TDMP) is guidance, training, and education in data literacy. To meet this need, our program has published the *Texas Data Management Framework* guide and regularly offers workshops and conferences related to data management best practices. This *Open Data Portal Resource Guide* is a part of the ongoing data literacy activities of the TDMP and provides formal training and education about the Texas Open Data Portal.

As Administrator of the Texas Open Data Portal (ODP), I am pleased to present the *Open Data Portal Resource Guide* as a resource for current and future participants of the ODP. While not exhaustive, this document serves as a guide for Texas public entities in establishing an open data governance framework, creating an open data inventory, and publishing open data in an efficient and standardized manner. The *Open Data Portal Resource Guide* will exist as a living document and will be modified as needed to reflect changes in open data policies and best practices.

Sincerely,

Tammi Powell

Administrator Texas Open Data Portal
Texas Department of Information Resources

Acknowledgements

This resource guide would not have been possible without the time and support of our editors. The Office of the Chief Data Officer would especially like to thank the Texas Department of Agriculture Food and Nutrition Division, who generously shared their knowledge and expertise in data management and analytics to help refine this guide. Their thorough reading and insightful suggestions are greatly appreciated. We would also like to thank Tyler Technologies for reviewing this document and suggesting additional content related to the Socrata platform. Finally, we want to acknowledge our colleagues at the Department of Information Resources for their support of this project and help editing this guide.

1. Open Data Overview

1.1 What is Open Data?

Open data refers to information that can be freely used, re-used, and redistributed by anyone. Such data should be available at no direct cost and accessible to a wide range of users, typically through internet downloads.

Open data is made accessible to the public in a format that is easily consumed to be read, filtered, sorted, manipulated, and combined with other data. Common file types of machine-readable data include, but are not limited to:

> *Machine-readable data is information in a format that can be processed by a computer.*

- .csv (Comma Separated Values)
- .xml (Extensible Markup Language)
- .tsv (Tab Separated Values)
- geospatial data (.kml, .kmz, .json, ESRI Shapefile)

The federal Open Government Initiative of 2013 made open and machine-readable data the new default for public information. Government entities are beginning to move to a centric model in which more public datasets are published in machine-readable formats on a central platform – an important transition as government entities of all levels hold vast amounts of data.

Effective September 1, 2019, the Texas Legislature, through Senate Bill (SB) 819, ushered in important advances for open data in the State of Texas. SB 819 amended the Texas Government Code, Chapter 2054, to require the establishment of a central open data repository, later named the Texas Open Data Portal (ODP). Under Section 2054.070, Texas Government Code, state agencies and political subdivisions are granted shared access to the repository, which offers an easy way to publish publicly accessible information. Furthermore, this section also requires each state agency to prioritize using the ODP and actively collaborate with the Texas Department of Information Resources (DIR) on publicly accessible data issues.

Excluded Data

> *Personally Identifiable Information (PII) and Protected Health Information (PHI) are protected by law and **are not** Open Data.*

In defining open data, it is important to note what types of information are excluded from this term. Health and personal information protected by law, such as Personally Identifiable Information (PII) and Protected Health Information (PHI) as mandated by the Privacy Act of 1974, are common examples of information that are not open data.

Other examples include law enforcement data in compliance with the Criminal Justice Information Services (CJIS), education data in compliance with the Family Educational Rights and Privacy Act (FERPA), or any other data that agencies and government entities have deemed private or sensitive in nature. Also excluded from open data is data that state agencies have submitted to the Office of the Attorney General (OAG) and received approval not to disclose due to its sensitive nature.

Federal law strictly prohibits the sharing of this information with the public and regulates how this data may be shared between government agencies.

Common examples of PII/PHI data include:

- Names and initials in any combination
- Birthdate
- Age
- Gender
- Home address
- Telephone number
- Personal e-mail address
- Drivers' license number
- Medical or health information
- Financial information (credit card/PCI, billing info, account info)
- Identification number (i.e. Social Security Number)
- Health information
- Nationality
- Physical characteristics
- Racial or ethnic origin
- Religious, philosophical, or political beliefs
- Biometric data
- Household information
- Consumer purchase or billing history
- Marital status
- Location (e.g. GPS) info (including that provided by mobile devices)
- Unique device identifiers (IP/ MAC addresses)

It is important to note that this is not an exhaustive list. For more information as well as additional resources on handling protected personal information, visit the U.S. Department of Justice's page about the Privacy Act of 1974: https://www.justice.gov/opcl/overview-privacy-act-1974-2015-edition

1.2 The Value of Open Data

Open data is a public resource that offers both governmental and private entities economic, performance, and social value. The economic value of open data is that it allows private business to better understand potential markets and integrate government data into new, more innovative products and services. Open data also adds performance value by increasing government transparency and accountability to its citizens. Greater openness of government improves the quality and efficiency of government services by allowing citizens and policy makers to make more informed data-driven decisions.

Providing greater visibility into how government manages resources in turn adds social value to open data by building stronger partnerships between citizens and their government. Having free and open access to information is the hallmark of a healthy

democracy and encourages greater participation by citizens who feel empowered to give feedback and suggest improvements on how their government is working. To summarize, the value of open data is that it offers:

- economic value by encouraging innovation in products and services,
- performance value by improving quality and efficiency of government programs, and
- social value by increasing citizen participation in government.

Reduction of Public Information Requests

> *The people, in delegating authority, do not give their public servants the right to decide what is good for the people to know and what is not good for them to know. The people insist on remaining informed so that they may retain control over the instruments they have created.*
> *- Texas Government Code, Chapter 552*

Another important value of open data to Texas governmental bodies is its ability to fulfill citizen public information requests. With the objective of creating a more open government that serves the information needs of an informed citizenry, the Texas Legislature (63R), 1973, passed the Public Information Act (PIA). The PIA gives citizens the right to inspect or make copies of open government records in the form of a public information request (PIR). Under the PIA, governmental bodies are generally required to release information to citizens unless the information is excepted from disclosure. Governmental bodies must respond to a PIR promptly, generally within 10 business days. Responding to a PIR, however, is both time and resource intensive. While citizens may be charged a fee for fulfilling a PIR, governmental bodies pay a higher cost of lost opportunity time to focus on other core business activities.

In 2017, the 85th Texas Legislature passed SB 79, to allow governmental agencies to redirect a PIR to information that is available on sites on the Internet, such as the ODP. Redirecting constituents to the ODP is a great economic and performance benefit to governmental bodies as it reduces the number of PIR's, thus saving time and resources.

For more information visit the Texas Office of the Attorney General's page about public information requests:
https://www.texasattorneygeneral.gov/sites/default/files/files/divisions/open-government/publicinfo_hb.pdf

Self-Service Model

Publishing open data on the ODP is not only a benefit to governmental bodies but is also of value to constituents. While citizens retain the right to request information directly from a governmental entity, filling out a formal public information request is no longer the only means of retrieving publicly available information. Citizens can now take advantage of open data published on the ODP to self-serve their information needs. In the same way that

redirecting a PIR to open sources of information saves governmental bodies time and resources, so too are citizens saved valuable time by accessing the same information at their convenience on the ODP.

Citizens are not the only constituents self-serving their information needs through the ODP. Employees of governmental entities also use the ODP to discover and access their own data to complete their work. Using the ODP to host internal open data for data sharing makes government more efficient by reducing acquisition costs, redundancy and overhead. In addition, researchers from higher education and public education likewise have access to the open data that they need to complete projects and initiatives that support findings for the overall community.

1.3 The Open Data Lifecycle

The Open Data Lifecycle describes the process by which open data is selected, prepared, published, and consumed by the public. As the purpose of open data is to allow citizens free and open access to publicly held information, every part of the Open Data Lifecycle works with that end goal in mind within a user-centered design approach. When selecting which data to publish on the ODP, for example, consider first data that will receive the most public information requests by constituents. Knowing how data will be ultimately consumed likewise drives decisions about how to best prepare data for citizens.

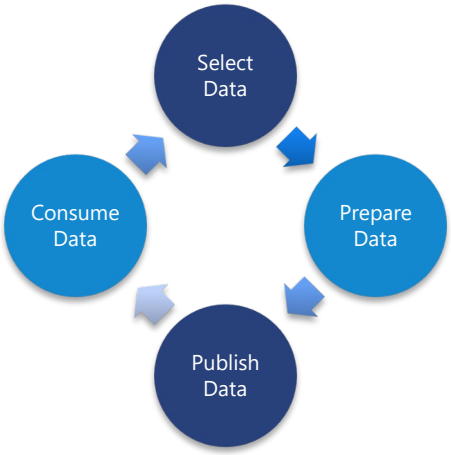


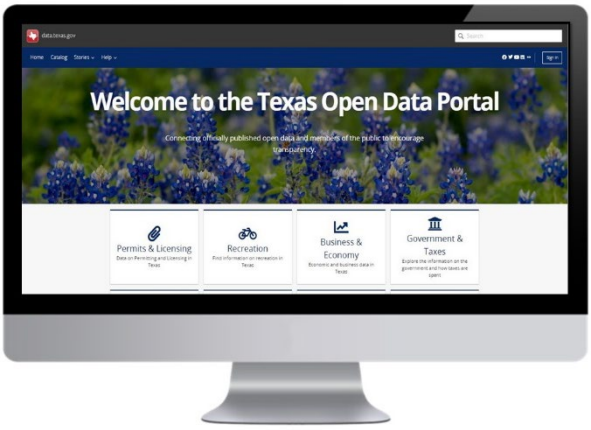
Figure 1: Open Data Life Cycle

Important to note is that publishing data is not simply making raw data available in an electronic format. Publishing also involves creating supporting documentation and metadata that gives constituents a greater understanding of what the data represents. At the end of the Open Data Life Cycle, citizens consume open data and create outcomes, such as measuring the performance of government programs or creating new products and services.

Tip: When selecting which data to publish to the Open Data Portal, consider first the information needs of end users in terms of fulfilling public information requests.

2. Open Data Portal

2.1 Open Data Portal



Background

The Texas Open Data Portal (ODP), located at https://data.texas.gov/, is the State's official website to access Texas open data. The portal is built on Socrata's *Open Data Portal* Platform, a software-as-a-solution (SaaS) developed specifically for government and recognized by Gartner as a predominant open data platform.

Differences with the Closed Data Portal

In addition to the ODP, agencies may also request their own instance of the Closed Data Portal. While the ODP is the official state of Texas central repository for all publicly accessible open data sets, the Closed Data Portal (CDP) is a private data sharing environment intended to host private or sensitive data. Unlike the ODP, access to data in the CDP is by invitation only. Use cases for the CDP include vertical data sharing between departments of the same agency as well as horizontal data sharing between various agencies collaborating on a common project or topic. While the ODP is managed by DIR, instances of the CDP are managed by individual agencies.

Feature	Open Data Portal	Closed Data Portal
Free of Charge to Agencies	X	X
Create, Manage, & Publish Datasets	X	X
Visualize Data with Charts & Maps	X	X
Data Automation	X	X
APIs for Datasets, Metadata, & Discovery	X	X
Private Data Sharing with Specific Audiences		X
Guest Role (share with external stakeholders)		X

Table 1: Comparison of Open and Closed Data Portals

2.2 Roles and Responsibilities

DIR Open Data Portal Team

The Department of Information Resources (DIR) Open Data Portal Team provides leadership and guidance on best practices for participating agencies as well as potential new open data publishers. The Open Data Portal Team is responsible for the day-to-day operations and management of the ODP, including:

- granting access and publishing rights to agencies,
- managing the dataset inventory,
- providing guidelines on open data governance,
- serving as a liaison between agency editors and the open data cloud provider, and
- facilitating open data review/design sessions with agencies to highlight best practices of open data selection, preparation, and publishing.

Open Data Portal Users Group (ODPUG)

To further facilitate collaboration and best practice sharing among ODP users with publishing rights, DIR re-established the **Open Data Portal Users Group (ODPUG) in 2020**. The group is comprised of state agency ODP users and/or representatives. The ODPUG meets regularly to discuss how to make improvements to the ODP, exchange success stories related to open data information sharing, and work with the Texas Data Program on open data policies and related issues. ODPUG agendas, decisions, and updates are shared with the Texas Enterprise Information Management (TEIM) data sharing community group.

Socrata

Socrata, a division of Tyler Technologies, is the vendor under contract with DIR that provides the off-the-shelf open data platform used for the Texas Open and Closed Data Portals. Socrata is responsible for:

- accommodating a range of structured dataset sizes and copies of previously published Texas reports and maps,
- providing a support process to log issues, track status, and request system enhancements,
- implementing multiple level security, including encryption of data in transit and at rest, and controls that provide assurance of availability, and
- participating in technical guidance and design sessions around data use cases.

Agencies

Agencies are the owners of all data uploaded to the ODP. As such, agencies are responsible for the content of the data published on the ODP, ensuring that data does not include content that either contains or is:

- objectionable
- unlawful
- in violation of a third party's copyright or intellectual rights
- personally identifiable or private/confidential information
- viruses or harmful files
- inaccurate or incomplete data

Note: For the purposes of this guide, the term "agency" is generally used to indicate a state agency or a state institution of higher education.

Warning: Agencies should only publish a **copy** of their data. The Open and Closed Data Portals are not the system of record of an agency's data.

3. Data Governance

3.1 Establishing an Open Data Governance Framework

A strong data governance framework is key to maintaining the availability, integrity, and security of an organization's information.

Data governance may be related to an agency's mission, apply to sensitive information, or can determine what type of information is available to the public.

It should be noted that an Open Data Governance Framework is part of a larger enterprise-wide data governance framework, which pertains to all data types that an organization manages. The Open Data Governance Framework follows similar methods and processes as the larger-scale governance model. The necessary details of your agency's Open Data Governance Framework will depend on the size, volume, and public interest in the public information it houses.

To establish this framework, your agency should first develop a structured charter or policy guide that identifies:

- the agency's goals and vision for open data,
- the roles and responsibilities of open data stakeholders and publishers,
- an open data inventory, and
- an approval/vetting process for publishing open data to the portal.

However, the formality of the charter or policies will vary by agency.

Open Data Stewards

Once an eligible agency has completed the onboarding process, it must identify its **Open Data Stewards**. Each steward is responsible for managing the publication of open data for their agency in their program area. Depending on the size and data assets of the agency, the number of stewards can range from one to ten or more employees.

Based on DIR's security recommendations, each steward with publishing rights should have a unique ODP login ID linked to their government entity-assigned email address. It is not recommended that publishers share ODP IDs or passwords, as this increases the risk of security breaches.

> *Each Open Data Steward should have a unique ODP ID. Sharing accounts among publishers is discouraged due to increased security risks.*

Open Data Stewards serve as key contributors in the following activities, which help improve the open data publishing process by:

- identifying open data to publish,
- ensuring data accuracy and quality,
- determining the frequency and method of updates, and

- working with agency leadership in reporting and analyzing open data consumption.

Agency leadership from both business and IT functions should collaborate to determine which employees are best suited to serve as Open Data Stewards, as well as the total number of stewards needed. Below is a list of suggested job titles that can be used to help guide the selection process:

- Database Administrator
- Computer Operations Technician
- Systems Analyst
- Business Analyst
- Programmer
- Computer Operations Specialist
- Web Administrator
- Planner
- Research Specialist
- Statistician/Economist
- Program Specialist/Manager
- Information Specialist

This list is not exhaustive but serves as a guideline for agencies. Depending on the level of technical complexity of open data publishing that the agency chooses to pursue, the skill level of its Open Data Stewards may vary, depending on the needs and demands of the agency.

It is critical that agencies understand their roles and responsibilities in publishing to the ODP. As *Section 2.2* outlines, agencies are solely responsible for the content, accuracy, and maintenance of their ODP datasets, and therefore bear all legal responsibility in the case of a security breach. Therefore, it is key for agencies (and especially users with publishing rights) to fully understand the characteristics of sensitive data that is protected by law. Examples of sensitive data are outlined in *Section 1: Open Data Overview*.

Warning: Agencies are responsible for implementing and following a data governance model that protects sensitive and confidential data from getting into the hands of unauthorized users.

3.2 Creating an Open Data Inventory

What is an Open Data Inventory?

As described in *Section 3.1*, each agency participating in the ODP should establish its own Open Data Governance Framework that is part of a greater enterprise-wide data governance practice. As a core part of this framework, an organization must determine what data is public or "open". Competing priorities, limited resources, and the often vast amounts of data available also require agencies to prioritize the publication of certain data that will hold more value for constituents and data consumers.

An Open Data Inventory is exactly that: an itemized list (typically stored electronically as a dataset itself) of available public datasets that the agency has published or may publish in the future.

Tip: Inventorying data assets is a continual process. Best practices dictate that data governance teams assess their open data inventory on a bi-annual basis to ensure that it reflects changes and updates to assets and metadata.

Open Data Inventory and Metadata

Your agency's datasets should also include applicable metadata. **Metadata**, or 'data about the data', is a set of data that provides information about each dataset. Metadata plays a vital role in the indexing and discoverability of each dataset and ensures that all personnel that handle or publish the data follow protocol.

> *Metadata is a set of data that provides information about each dataset. It makes finding and working with instances of data easier.*

Although the ODP currently only requires the fields of Title, Category Tile, and Agency, it is beneficial for agencies to collect more detailed information about each dataset when conducting an open data inventory. The more metadata that can be documented on published datasets, the easier it is for constituents to find and understand the context surrounding published data.

Other than giving basic context of each dataset, the inventory should also provide information about:

- file type,
- size,
- quality, owner,
- update frequency, and
- security/privacy constraints of the data.

TESTIMONIAL

"The Texas Open Data Portal's proactive approach to delivering data to stakeholders and the public allows them real-time access when they need it most.

Fast and efficient access fuels the support and understanding of our mission — feeding the hungry and promoting healthy lifestyles. In addition, this self-service option reduces the amount of staff time used to pull data for Public Information Requests, thereby affording more time for staff to ensure compliance and accountability in the administration of 12 federal nutrition programs in Texas.

The Texas Open Data Portal has the potential to become a vital tool in helping TDA Food and Nutrition meet its primary objectives of providing customer service excellence and achieving high professional standards."

— Angela Olige, Assistant Commissioner, Texas Department of Agriculture

4. Data Preparation

4.1 Data Assessment and Selection

As your agency begins to build its own Open Data Inventory, agency personnel must decide which datasets offer the most value to the public and any consumers of the data. Approaching data selection from a user-centered design model is key to the success of an organization's open data program. Data stewards should accordingly select data for its value, quality, completeness, and appropriateness according to constituent information needs.

Guiding Questions for Data Assessment

The following guiding questions will help identify and prioritize open data to create an Open Data Inventory for publication.

- What information is frequently requested by the public, via the Public Information Request process?
- What data does the agency or program areas use for key performance indicators (either State Auditor's Office or internal measurements)?
- What data populates monthly, quarterly, or annual agency reporting requirements?
- What information is reported to local, state, or federal agencies?
- What information is shared with other agencies?
- What information is shared with external stakeholders?
- What kinds of open data are peer agencies across the country publishing?

In addition to the questions above, other factors that must be considered include:

- the quality of the data,
- how often new data will be available (update frequency),
- file format (machine-readable),
- the ability to automate publication after the initial dataset is made public (see *Section 5.3*), and
- accessibility to data source systems to facilitate automation.

4.2 Data Quality

For data to be useful, it must be reliable, trustworthy and of high quality. Prior to publishing a dataset on the ODP, data stewards should first complete several steps to ensure a high quality, interoperable and usable product for data consumers.

The core elements of data quality are:

- Completeness
- Uniqueness
- Timeliness
- Validity
- Accuracy
- Consistency

> *If data does not meet quality standards, analysis will be suspect and invalid.*

Data Quality Checkpoints

By considering the following data quality checkpoints organizations maximize the discoverability, analysis, reuse and understanding of their open data assets.

- Datasets should be formatted in a **machine-readable format** that allows for the use and reuse of data. Comma Separated Values (CSV) are the standard format for publication.
- Use **raw data** instead of calculated columns to make data more flexible and consumable in a variety of ways.
- Use vertical rather than horizontal **orientation of data** to make datasets more human readable and easier to create visualizations.
- Use **column names** that reflect natural language instead of the source data system.
- Include meaningful data points instead of having **blank cells** or a value of "N/A".
- Ensure that **data is consistent** (e.g. avoid storing both text and numeric values in a numeric field).
- If **coded fields** are included, add a descriptive column of the text equivalent for readability.
- Trim text fields of **white space.**
- **Numeric data** that represents money should either have two decimal places or none. To ensure ease of use, do not include currency symbols or commas.
- **Measurements** may include a varying number of decimal places. However, do not use commas or parenthesis.
- Where **leading zeroes** are meaningful (e.g. identification numbers), store data as text to ensure that values are not truncated.
- **Trim trailing spaces** so that operations which require an exact match are not affected.
- Use the appropriate **data types** that reflect how the data will be used.
- Store **dates** in the MM/DD/YYYY format to make historical analysis easier.
- Ensure that data within the same **field is consistent** in terms of data type, format and expected values.

4.3 Data Privacy

Data privacy is the part of data security concerned with data handling ethics. While data security protects data from internal and externa attackers, data privacy governs how sensitive data is shared, collected, stored, and regulated. *Section 1.1* gave examples of sensitive data that is excluded from open data sets. This section outlines best practices in handling sensitive data when determining what data to publish to the ODP.

Guiding Questions for Data Privacy

- Does publishing the data violate any laws, rules, or regulations?
- What are the intended and unintended effects of publishing the data?
- If the data does not contain personal identifiable information, can it still be combined with other public information to reveal sensitive details?
- Does aggregated and anonymized data have the potential to compromise internal processes, such as arrest records that expose where police have concentrated their efforts?
- Does your organization own and have the right to publish the data? Was the data collected and compiled by a third party?

Tip: Always consult with your General Counsel regarding the rules and regulations specific to your agency's data publishing policies and practices.

TESTIMONIAL

"The Texas Open Data Portal played a key role in modernizing our website. Using Socrata's tools, we were able to increase our efficiency by automating the transfer of data to our website. The portal also boosted our Open Data presence and created new opportunities for collaboration with other state agencies."

— Robert Wood, Associate Deputy Comptroller, Operations and Support

5. Publishing Open Data

5.1 Creating Metadata for a Dataset

A critical first step in publishing data on the Texas Open Data Portal (ODP) is creating metadata that describes the dataset and columns. Providing quality metadata is important to users because it creates access points to find, gives context to understand, and clarifies meaning to interpret the data. Below are the metadata fields for datasets and columns visible on each dataset's main metadata page.

Dataset Level Metadata

Dataset level metadata describes the dataset as a whole and helps users discover datasets in the ODP's data catalog. Dataset level metadata gives general information about the title, subject, and source of the data.

Metadata Field	Justification	Guidelines
Title (Required)	Title helps discover and select datasets as well as differentiate between similar datasets.	Human-readable name of the dataset in plain language that includes enough detail to facilitate searching. Free of acronyms.
Description	Description helps discover and select datasets as well as differentiate between similar datasets.	Human-readable description with sufficient detail to enable the non-technical user to quickly understand whether the dataset is relevant to their search.
Row label	Only if applicable.	Human-readable description of what each row in the dataset represents.
Tags / Keywords	Tags provide a means to include technical language, secondary categories, and acronyms.	Generally single words which help users discover the data and include terms that would be used by technical and non-technical users.
Public License Type	A license reduces legal uncertainty for data consumers or users.	Choose licensing terms for people who wish to use this dataset.
Data Provided By	Users can filter datasets by publishing agency.	Individual or organization that provided data.
Source Link	Only if applicable.	If applicable, the link to the source of the data.

Metadata Field	Justification	Guidelines
Email Address	The email address that users can use to ask questions about the dataset.	Who manages the data and is responsible for making changes to the data? Who understands what the dataset includes and can answer questions about it?
Agency (Required)	Users can browse the data catalog and filter datasets by the publishing agency.	Choose the agency responsible for publishing the data.
Category Tile (Required)	Category provides a distinct navigation method and groups similar datasets together regardless of source.	The category of the dataset identified by the list of possible values. If a data set can fall into multiple categories, select the one which is most significant.

Table 2: Dataset Metadata Fields

Column Level Metadata

Like a data dictionary, metadata about data columns or fields helps users better interpret and use datasets on the ODP. For users who frequently use and analyze data, column level metadata is important as it gives context and meaning to the data. By creating quality column level metadata, agencies allow their constituents greater access to the information they seek.

Metadata Field	Guidelines
Column Name (Required)	Human-readable name of a column in plain language that is free of jargon and acronyms.
Column Description	Human-readable description with sufficient detail to enable the user to interpret data in a column.

Table 3: Column Metadata Fields

Tip: When creating metadata consider first what terms make the most sense to end users. Using language that most closely matches user search terms and expectations will help users better find, identify, select, and obtain the information they seek.

Learn more on Socrata's page about metadata best practices: https://support.socrata.com/hc/en-us/articles/115008609707-Best-Practices-for-Metadata-Management

5.2 Manual Publishing

Manual publishing is performed by users with publishing rights, called data editors, and is the simplest way to upload a dataset to the ODP. Using this method, an editor can upload an individual file (.csv, .xls, .xlsx, .tsv) or create a dataset from scratch. Editors can also make necessary edits before publishing. Once marked as "public", each dataset can be manually or automatically updated as new data becomes available.

Learn more about how to manually publish data on Socrata's data management page: https://support.socrata.com/hc/en-us/articles/115016067067-Using-the-Socrata-Data-Management-Experience

5.3 Automated Publishing

Automated publishing is the process of updating data programmatically, rather than manually. The advantages of using automated publishing is that updates can be scheduled without the delay of relying on an individual to manually refresh data. This section highlights two popular modes of automated publishing on the ODP using DataSync and the Open Data API.

DataSync

The **Socrata DataSync** tool is a free and powerful publishing tool that lets users schedule and automate data updates and upload large data files to the ODP. The program requires Java version 8 or higher, and can be downloaded for free from https://github.com/socrata/datasync/releases.

When you have downloaded DataSync and opened the program, the window will look like what is displayed in *Figure 2* below.

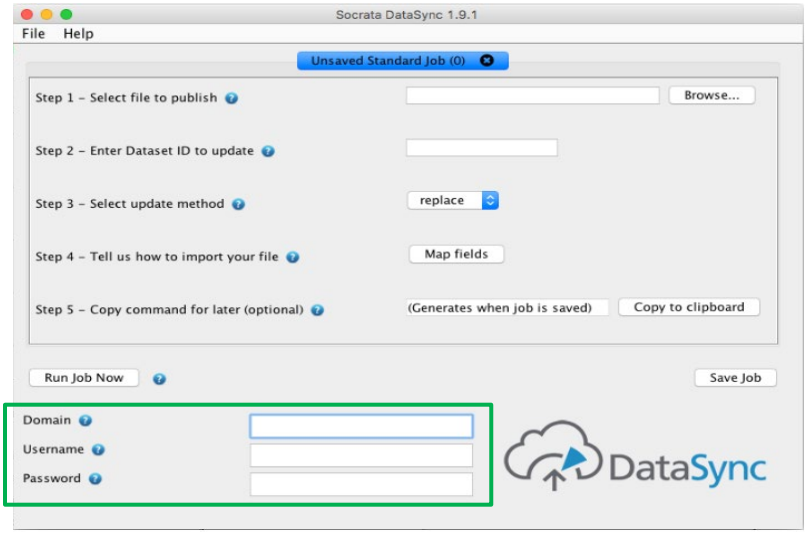


Figure 2: DataSync Automated Publishing

To begin the automated publishing process, follow these steps:

Step 1: Select the file to publish (.csv/.tsv files only).

Step 2: Enter Dataset ID to update (the 8 character code separated by a dash at the end of a dataset's URL).

Step 3: Select update method (option to replace, upsert, append, or delete). Visit the following page for more details: https://support.socrata.com/hc/en-us/articles/210850597-Getting-Started-with-DataSync#datasync-prefs

Step 4: Indicate how to import the file (to ensure map fields are correctly coded).

Step 5: Copy command for later use (optional, use to copy and save a command to run from the command line or to schedule future jobs).

To complete the Authentication Details Section (highlighted in the image above), enter the following information:

Domain: https://data.texas.gov

Username/Password: Socrata Open Data Portal login credentials.

App Token: A user-specific code to use DataSync. Visit the following page for instructions on generating app tokens: https://support.socrata.com/hc/en-us/articles/210138558

Socrata Open Data API

The ODP also has tools that allow agencies to keep their own website data synced with datasets on the portal. Each uploaded dataset comes with a built-in system called the **Socrata Open Data API** that allows websites, as well as many other types of applications, to communicate directly with the Socrata platform via code.

> *API (Application Programming Interface) is a predefined way to allow one computer program – such as a website – to communicate with another (like the Socrata platform).*

This process may take a little extra time for a developer to initially set up, but the benefit is that your website or application will automatically refresh as the ODP datasets are updated.

A few situations where an agency might consider using the API include:

- updating data on the ODP and then updating the same data on an agency's website as a separate process,
- updating a dataset on an agency's website at regular intervals,
- adding more context to a dataset than is possible on the Socrata platform,
- presenting a dataset in a manner that is not readily available using the ODP, or
- due to limited time and/or resources, updating open data on a consistent, manual basis is not feasible.

* Note: Automated data publishing should be implemented with the full compliance of an agency's data governance process.

For more information visit Socrata's developer page at https://dev.socrata.com/ for a better understanding of the API's capabilities, as well as technical information on how to get started.

5.4 Publishing Process

Whether you are using the manual or automated method to publish datasets to the ODP, the publishing process should generally follow the cycle illustrated in Figure 3.

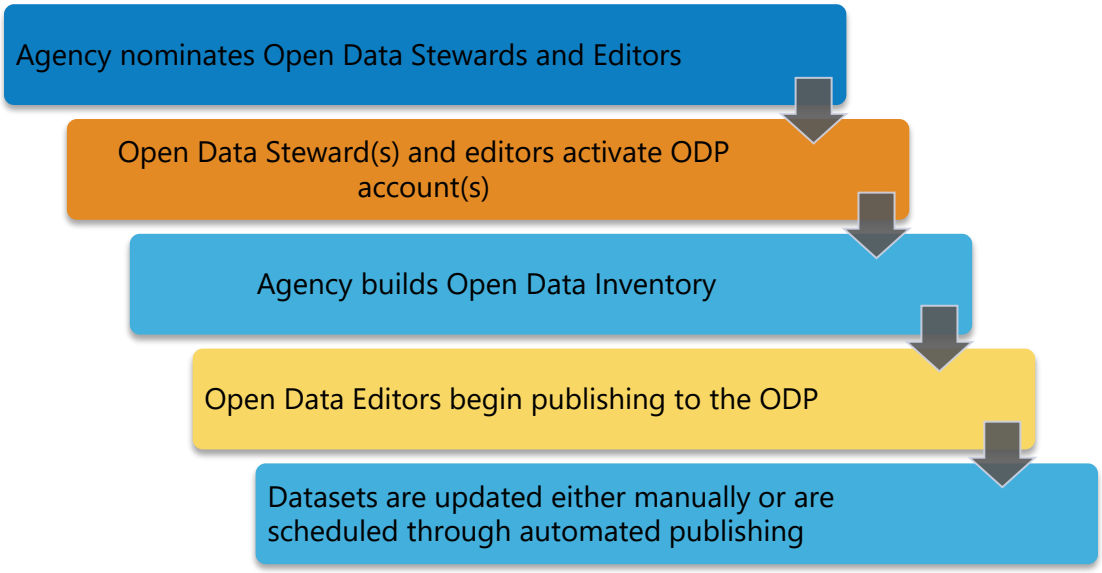


Figure 3: The Publishing Process

Private vs. Public Data

The ODP has another useful function for data editors – the ability to mark a dataset as "private" before it is marked "public". When an editor uploads a new dataset, it remains in private mode until the "make dataset public" option is selected. While in private mode, only data editors and ODP administrators can view or make changes to the dataset. This enables several Open Data Stewards within the same agency to review a dataset before publication, ensuring that data quality, accuracy, privacy and/or security concerns are addressed.

Data editors can also share editing permissions with colleagues within the same agency, to allow for a vetting process of datasets before they are published to the public. The table below explains the types of users, with permissions indicated with an "X":

Domain-Level Actions	Admin	Editor	Designer	Viewer	Public
Add, edit, and delete user accounts	X				
Manage homepage content, including story banner and featured datasets	X		X		
View site-wide analytics	X	X	X	X	
Manage metadata settings: categories, custom fields	X		X		
Upload and own datasets	X	X	X		
Publish a dataset draft	X	X	X		
Delete a dataset	X	X	X		
View open data catalog list of assets	X	X	X	X	X
Discover ALL assets - public and private - across the site	X				

Table 4: Open Data Portal Users and Permissions

Read more about publishing on the ODP on Socrata's data publishing page: https://support.socrata.com/hc/en-us/categories/360000024687-Data-Publishing.

6. Data Consumption

In addition to the data publishing methods described above, the Open Data Portal (ODP) also provides filtering and graphic data visualization capabilities for editors and viewers alike. Each of the features described below update automatically as the underlying dataset is refreshed. This section provides general guidance in using these functions, but more detailed information, tutorials, and videos can be found on Socrata's page on creating visualizations: https://support.socrata.com/hc/en-us/articles/115000813847-Creating-a-Visualization-in-the-Visualization-Canvas

6.1 Filtering and Sorting

Data consumers have the option of either exporting raw datasets into their own data analytics software or filtering and sorting data directly in the ODP. Users with a Socrata account can then save these filtered views. Some ways which users can filter datasets include:

- filter columns by group,
- conditional formatting,
- sorting, and
- roll-ups and drill downs.

6.2 Charts, Graphs, and Maps

The basic graphing, charting, and mapping capabilities can be leveraged by any type of user (viewer, editor, administrator, etc.). Editors can also use these features and publish them as "official views" so that they appear in the data catalog. If a constituent has created a Socrata account and is logged in, he or she can create and save filtered views, maps, or charts of official datasets but cannot publish original data.

Charts, graphs, and map features available on the ODP include:

- calendar mapping (column data must be formatted as date),
- mapping (column with address, zip code, or city must be formatted as location), and
- charts, depending on data type (column, bar, pie, donut, line, area, timeline, bubble, and tree maps).

Some of these functions are displayed in the graphics below.

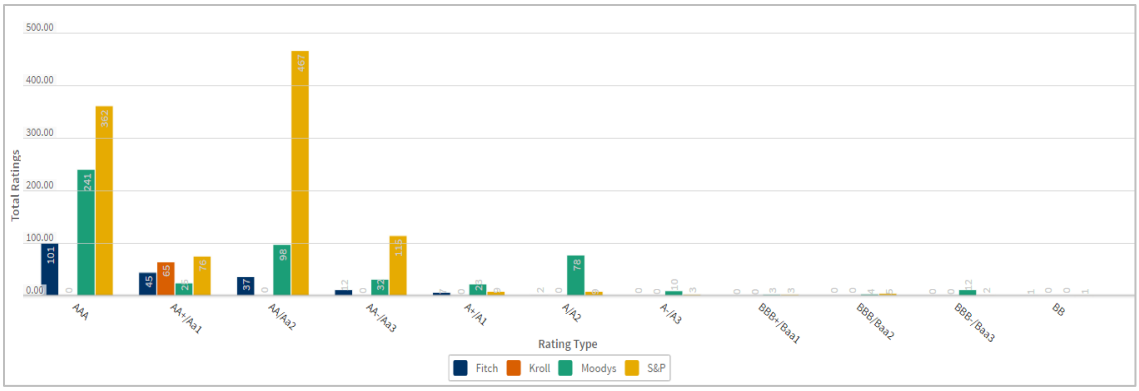


Figure 4: Clustered Bar Graph

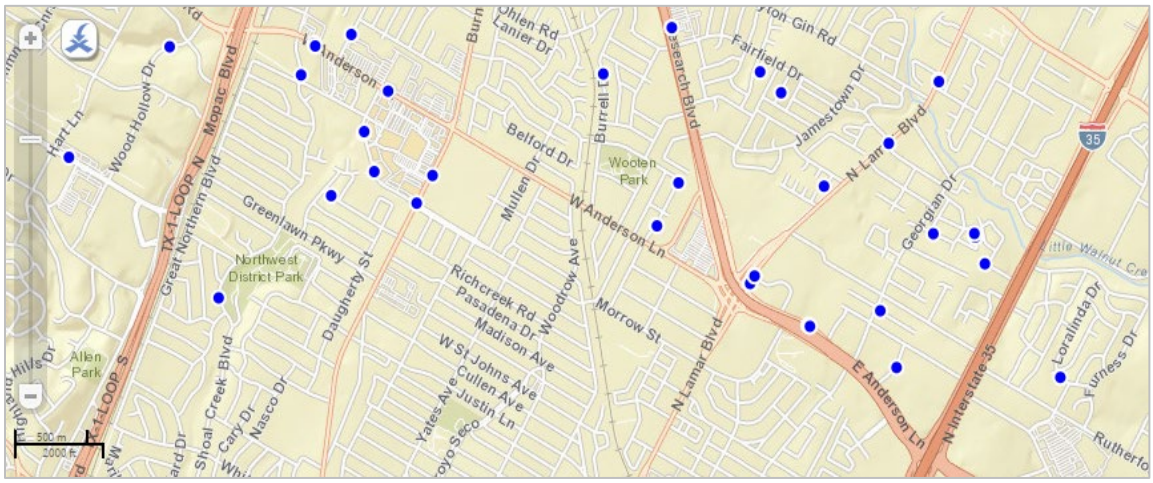


Figure 5: Point Map by Location

Data Lens

The **Socrata Data Lens** tool allows ODP editors to create data visualizations that the public can easily engage and interact with by customizing how data is presented in the catalog. Datasets with geographic information, such as address, city, or zip code can be mapped both by point location and by aggregate comparisons by county, Council of Government, Senate and House District.

Other functions include search bars by column and bar graphs. Figures 6 and 7 show examples of Data Lens visualizations; these engaging visuals can also be exported to another website via API or downloaded as a .png file.

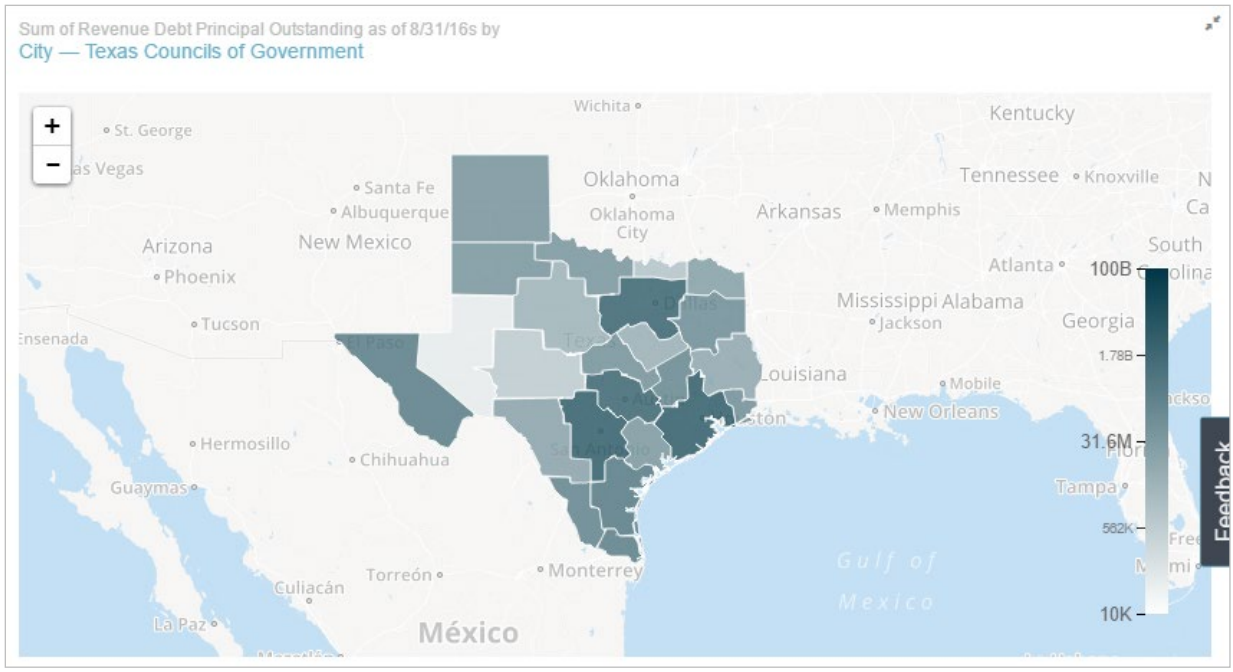


Figure 6: Data Lens Heat Map

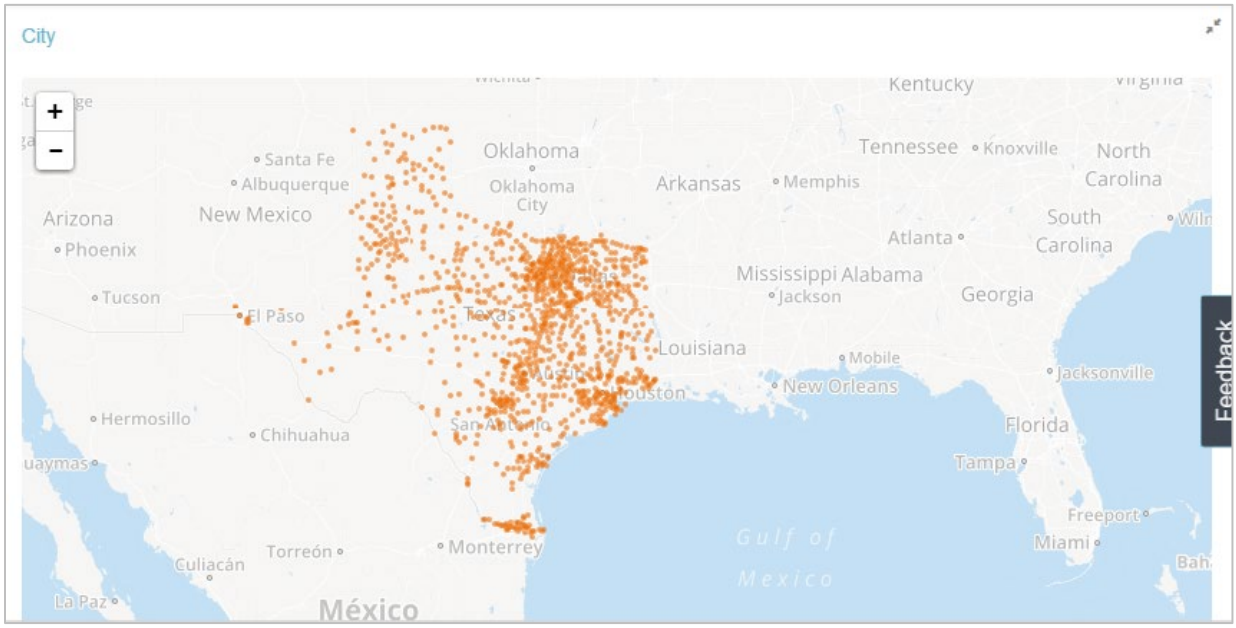


Figure 7: Data Lens Point Map

6.3 Data Stories

Agencies can also create their own webpages on the ODP called data stories. Data stories are multimedia pages that combine text, visualizations, and images to tell a story about an agency's data and programs. While individual data sets allow users access to raw data, data

stories provide greater context and meaning to what the data represents and why it is important to consumers. Data stories are outreach tools that highlight and bring recognition to the important work agencies do for the public.

Users have the following options for adding and editing rich media content in data stories:

- Socrata Visualization: create or insert an existing chart, map, or table.
- Socrata Story Tile: insert a preview of another story.
- Socrata Open Performance Measure Tile: insert a preview of an Open Performance measure.
- YouTube: add a video from a YouTube link.
- Image: add an image from a computer file or from Socrata's Getty images library.
- HTML Embed: insert an embed code to include external web content.

Below is an example of a data story published on the ODP that includes an embedded map, a performance measure, as well as explanatory text that further gives context to the various data available through this program. Stories are also dynamic webpages. As datasets are updated in the ODP, the visualizations and performance measures created from the dataset are also refreshed in real-time, including those embedded in a story page.

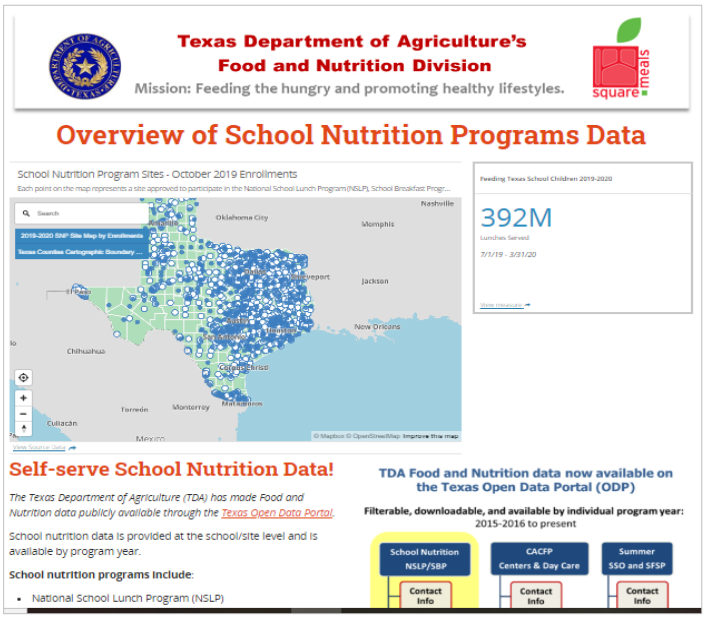


Figure 8: Texas Department of Agriculture Data Story

Learn more on Socrata's page about creating data stories: https://support.socrata.com/hc/en-us/articles/218691088-Creating-a-Story

Read more articles related to data visualization and exploration tools on the ODP on Socrata's data visualization page: https://support.socrata.com/hc/en-us/categories/360000027108-Data-Visualization-and-Exploration

7. Checklist for Getting Started on the Open Data Portal

It is the role of the DIR Administrator of the Open Data Portal (ODP) to onboard new customers as well as support existing ODP users in their open data journey. The following checklist is used to onboard new users and ensure their continued success in the Texas Data Management Program.

The ODP Administrator is a resource and supports users in each of the following steps:

- ☐ Prospective agency agrees to the **terms and conditions** of the ODP.
- ☐ Agency identifies **data stewards** who create a **data inventory** of potential open data assets and may also act as data editors.
- ☐ Agency submits the names of **data editors** to the ODP Administrator, who will register authorized staff on the Socrata platform.
- ☐ Data editors register for the **Socrata Open Data Portal Education Plan** outlined in *Section 8* to learn how to use the Socrata platform.
- ☐ Agency open data governance team **selects and prepares data for publication** on the ODP. The DIR Administrator of the ODP is available as a resource to assist with identifying data for ODP publication.
- ☐ Data editors practice publishing and other skills on the **sandbox instance** of the ODP at https://texas-test.demo.socrata.com/login.
- ☐ Data editors work with data stewards to **create metadata and publish open data** to the ODP.
- ☐ Data stewards and data editors join the **Open Data Portal Users Group (ODPUG)** to network, share and learn from other users.
- ☐ ODP users and members of the open data governance team attend **trainings, webinars and conferences** hosted by the Office of the Chief Data Officer, including the annual **Open Data Workshop and Data Forum.**

8. Socrata Open Data Portal Education Plan

Socrata has a library of on-demand and live web courses and learning paths designed to help data stewards and editors learn the platform and publish their data. Users must first log in to access course material. To access the site and set up an account, visit Socrata's Online Training Academy at http://learn.socrata.com/?access_code=texas_dir using the code **texas_dir**.

Recommended Course Order for New Users

1. **Introduction to the World of Data on Socrata**
   http://learn.socrata.com/introduction-to-the-world-of-data-on-socrata-ondemand
2. **Prepare Your Data for Socrata**
   http://learn.socrata.com/prepare-your-data-for-socrata-ondemand
3. **Create and Manage Your Socrata Dataset**
   http://learn.socrata.com/create-and-manage-your-socrata-dataset-ondemand
4. **Explore Data with Charts on Socrata**
   http://learn.socrata.com/explore-data-with-charts-on-socrata-ondemand
5. **Map Your Data on Socrata**
   http://learn.socrata.com/map-your-data-on-socrata-ondemand
6. **Tell a Story with Socrata Perspectives**
   http://learn.socrata.com/tell-a-story-with-socrata-perspectives-ondemand

Foundational Courses

- **The Story of Open Data**
  http://learn.socrata.com/the-story-of-open-data-ondemand
  Where did the open data movement start and why? How does that fit into your day-to-day? What are data, datasets, and how are they different from databases? This course answers these questions and more.
- **Map Out Your Digital Center of Excellence Strategy**
  http://learn.socrata.com/map-out-your-digital-center-of-excellence-strategy
  A Digital Center of Excellence (DCoE) acts as a hub of best practices, policies, data standards, education, and support to scale the data program across divisions and departments. Learn how to get started with yours.
- **Developing Your Data Governance Policies and Standards**
  http://learn.socrata.com/developing-your-data-governance-policies-and-standards
  - Identify elements from the data governance policies of other organizations that would work in your own organization.
  - Identify the factors that influence standards for metadata and data quality.
  - Understand the elements of effective data sharing and collaboration.
  - Draft a data governance policy.

9. Additional Resources and Links

Socrata Support

Socrata Knowledge Base Articles: https://support.socrata.com/hc/en-us

Socrata Support Ticket Submission: https://support.socrata.com/hc/en-us/requests/new

Socrata Developer Portal: https://dev.socrata.com/

External Open Data Portals

Federal Open Data Portal: https://www.data.gov/

City of Austin Open Data Portal: https://data.austintexas.gov/

City of Dallas Open Data Portal: https://www.dallasopendata.com/

10. Glossary of Terms

API – stands for Application Program Interface. API is programming code that allows for two distinct applications to communicate with each other in an automated way.

Closed Data Portal – a private data sharing environment intended to host private or sensitive data. Unlike the Open Data Portal (ODP), access to data in the Closed Data Portal is by invitation only.

Data – information that has been translated into binary components that computers can move, store and process.

Dataset – a collection of data, usually in the form of tabular data, that is published on the ODP.

DataSync – the Socrata proprietary tool that has been developed to automate the data publishing process for ODP users.

Data Steward – the person or persons responsible for management and oversight of an agency's data. Data stewards provide a vital role is ensuring data quality, availability, and consistency on behalf of the data owner.

Editor – an ODP user with publishing rights.

Machine-readable – data in a standard format that can be easily recognized and processed by computers.

Metadata – 'data about the data' that provides the contextual understanding about each dataset.

Open Data – data that refers to information that can be freely used, re-used, and redistributed by anyone.

Open Data Governance Framework - a part of the larger data governance framework of people, process, and technology to manage an agency's information. The framework for open data is the key to maintaining the availability, integrity, and security of an organization's information.

Open Data Inventory – an itemized list (typically stored electronically as a dataset itself) of available public datasets that the agency has published or may publish in the future.

Texas Data Management Framework (TDMF) – the methodology of core principles for developing individual agency Enterprise Information Management programs.

Texas Public Information Act - the policy of the State that, unless expressed otherwise by law, guarantees citizens complete information about the affairs of government and the official acts of public officials and employees.

11. Appendix: Open Data Legislation in Texas

Texas Public Information Act – Title 5, Open Government, Ethics, Chapter 552 – Public Information (September 1, 1993) Set forth the policy of the State that, unless expressed otherwise by law, guarantees citizens complete information about the affairs of government and the official acts of public officials and employees.
http://www.statutes.legis.state.tx.us/Docs/GV/htm/GV.552.htm

Open Data – SB 701 82(R) (September 1, 2011) Required state agencies to post high-value data sets on a generally accessible agency website. High-value data sets include, but are not limited to, data that is critical to the financial and programmatic function of state agencies.
http://www.capitol.state.tx.us/tlodocs/82R/billtext/pdf/SB00701F.pdf#navpanes=0

Open Data – SB 279 83(R) (September 1, 2013) Required state agencies to send the Department of Information Resources (DIR) a description and link to their high value data sets.
http://www.capitol.state.tx.us/tlodocs/83R/billtext/pdf/SB00279F.pdf#navpanes=0

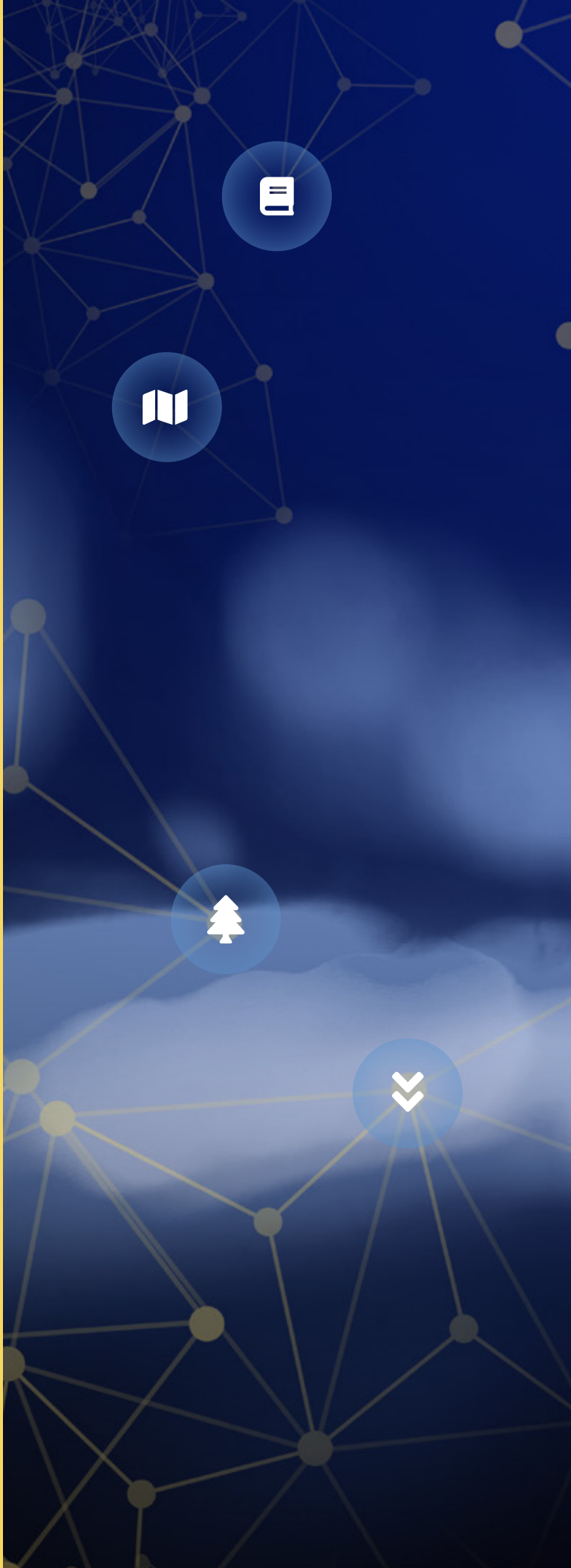
Statewide Data Coordinator – HB 1912 84(R) (September 1, 2015) Created position to collaboratively develop data policies, standards, and best practices and to improve data governance and integrity statewide.
http://www.capitol.state.tx.us/tlodocs/84R/billtext/pdf/HB01912F.pdf#navpanes=0

Interagency Data Transparency Commission – SB 1844 84 (R) (September 1, 2015) Created the commission to study and review the current public data structure, classification, sharing, and reporting protocols for state agencies.
http://www.capitol.state.tx.us/tlodocs/84R/billtext/pdf/SB01844F.pdf#navpanes=0

Chief Data Officer and Texas Open Data Portal – SB 819 86(R) (September 1, 2019) Established the position of Chief Data Officer (CDO) for the State of Texas and established the Open Data Portal (ODP) as the official central repository for publicly accessible electronic data.
https://capitol.texas.gov/tlodocs/86R/billtext/pdf/SB00819F.pdf#navpanes=0

Notes

Notes



Texas Department of Information Resources
300 W. 15th Street, Suite 1300, Austin, TX 78701

Transforming How Texas Government Serves Texans
dir.texas.gov | #DIRisIT | @TexasDIR